

DIKUSA

Teilprojekt: Aus, nach und innerhalb von Deutschland migrierte Frauen – Aufbau einer erfahrungsgeschichtlichen Wissensbasis

Datenmanagementplan

Erstellt durch: Francesca Weil, Nicole Husemann, Chris Weinhold, Sophie Seeliger, Dirk Goldhahn

Datum: 21.11.2024

Version: 1.0

1. Administrative Angaben

Projekttitle: Aus, nach und innerhalb von Deutschland migrierte Frauen – Aufbau einer erfahrungsgeschichtlichen Wissensbasis, Teil des Projekts: Vernetzung digitaler Kulturdaten in Sachsen – Aufbau einer technischen Infrastruktur für die Forschung zu Mobilität, Migration und Transformation von Orten, Personen und Artefakten (in zeitlicher und räumlicher Perspektive) – DIKUSA

Antragsnummer (SAB): 100611724

Kontonummer (SAB): 3000954790; Bewilligung vom 24.02.2022

Kundennummer (SAB): 2002094846

Clusternummer (SAB): 4192

Fördergeber: Landesmittel des Freistaats Sachsen, Titelgruppe 70. Die Zuwendung stammt aus Steuermitteln. Diese Steuermittel werden auf der Grundlage des vom Sächsischen Landtag beschlossenen Haushaltes zur Verfügung gestellt. Details siehe Projektskizze, Vollerträge der 6 Partner und 6 Zuwendungsbescheide der SAB.

Zeitraum: 01.02.2022-31.12.2025 (Teilprojekte können beim Beginn leicht abweichen)

Mitarbeitende/Datenurheber: Francesca Weil, Nicole Husemann, Chris Weinhold, Sophie Seeliger

Koordinierung: KompetenzwerkD an der SAW - Dr. Dirk Goldhahn, Mag. Peter Mühleder, PD Dr. Franziska Naether

Antragsteller: Sächsische Akademie der Wissenschaften zu Leipzig, vertreten durch den Präsidenten Prof. Dr. Hans Wiesmeth und den Generalsekretär Dr. Christian Winter

Kontakt: kompetenzwerkd@saw-leipzig.de

Kurzbeschreibung/Abstract Gesamtvorhaben

Im Forschungsalltag stellt die niedrighschwellige digitale Wissenserfassung und -integration die außeruniversitären geisteswissenschaftlichen Forschungseinrichtungen Sachsens vor unterschiedlich große Herausforderungen. Ziel dieses Verbundvorhabens ist es, eine entsprechende technische Infrastruktur zu entwickeln und die Kompetenz der Häuser für die digitale Aufnahme von Archivmaterial und Objektdaten, deren Verlinkung, Visualisierungen sowie den Abgleich mit Normdatensätzen zu ermöglichen. Dafür wird zusammen mit dem KompetenzwerkD und der SLUB eine technische Basis bestehend aus mehreren Komponenten erstellt: eine zentrale Wissensbasis mit Referenzierungs- und Reconciliation-Diensten als Datenhub, eine „Toolbox“ mit nachnutzbaren Softwarekomponenten für die Datenarbeit und Dienste für die Visualisierung auf historischem Kartenmaterial. Dies ermöglicht zugleich eine attraktive Außendarstellung der Teilprojekte. Die technische In-

frastruktur wird im Rahmen von sechs Teilprojekten aus dem Bereich der Geisteswissenschaften und der Digital Humanities entwickelt, die Kulturdaten über Orte, Personen und Artefakte in der zeitlichen Perspektive seit dem Mittelalter bis zur Gegenwart in spezifischen Fragestellungen in den Blick nehmen.

Erster Pressebericht: <https://www.saw-leipzig.de/de/aktuelles/neues-verbundprojekt>

Kurzbeschreibung/Abstract Teilprojekt

In diesem Projekt sollen Ursachen und Motive von Migration inklusive Flucht in den erfahrungsgeschichtlichen Blick genommen werden. Dafür wird eine Forschungsplattform umgesetzt, die es ermöglicht, Daten zu den klassischen drei Phasen der Migration zu erfassen: 1. die Situation der Frauen in den jeweiligen Herkunftsländern, 2. die Erfahrungen während der Migration und 3. die Integration in den Ankunftsändern inklusive des Verhaltens der Aufnahmegesellschaften.

Vor dem Hintergrund, dass Verfolgung, Flucht und Integration komplexe Phänomene darstellen, werden neben frauentypischen Migrationsmotiven und -erfahrungen sowie frauentypischen Integrationserfahrungen persönliche, politische, strukturelle, gesellschaftliche, religiöse, emotionale und rassistische Aspekte und deren mögliche Verknüpfung und Widersprüchlichkeit herausgearbeitet. Das Schlüsselthema einer gender-bewussten Migrationsforschung, das in diesem Projekt im Mittelpunkt steht, ist die Frage nach den emanzipatorischen Auswirkungen, die sich für die Frauen aus der Migration ergeben. Zu untersuchen gilt es, inwiefern und unter welchen Bedingungen die Migration für Frauen berufliche und andere Chancen eröffnete, die sie in ihren Heimatländern in diesem Maße nicht gehabt hätten, und die ihre emanzipatorische Entwicklung vorantrieben, bzw. welche Barrieren und Grenzen dem entgegenstanden. Darüber hinaus muss der Frage nachgegangen werden, ob sich migrierte Frauen in ihren Ankunftsändern auch mit ihrer Situation als Frau viel bewusster auseinandersetzen als vorher und mehr Selbstbestimmung entwickelten.

Es wurde zu Recht erkannt, dass die Probleme der Frauen nicht nur mit Besonderheiten der Frauenmigration zu tun haben, sondern vor allem mit der Bereitschaft der Aufnahmegesellschaften, unterschiedliche Kulturen als gleichwertig anzusehen. Deshalb wird sich auch mit Ressourcen und Potenzialen speziell von Frauen beschäftigt, die sie befähigen zu migrieren, ein neues Leben zu organisieren und eine neue Identität zu entwickeln, ohne sich vollkommen zu assimilieren, d. h. ohne sich komplett anzupassen oder ihre Herkunft und Kultur zu verleugnen. Hierbei spielt auch die Inklusion bzw. die Akkulturation eine wesentliche Rolle. Letztere fokussiert auf die Annahme von kulturellen Angeboten der Ankunfts-gesellschaft, auf die Annäherung an Identitätsmuster der umgebenden Mehrheitskultur bei Bewahrung bestimmter eigener Traditionen. Insgesamt ist damit ein beiderseitiger Prozess gemeint, denn auch die Mehrheitsgesellschaft wandelt sich unter Rück- und Einwirkung der einwandernden Minderheiten.

Ziel des Projektes ist es, eine Infrastruktur zu entwickeln, welche die Zusammenstellung strukturierter digitaler Daten über die Migrantinnen erlaubt. Dies umfasst ihre Lebensdaten sowie weitere damit in Verbindung stehende Informationen mit dem Ziel, diese in nachhaltig nachnutzbarer Form bereitzustellen. Ein zentrales Anliegen dieses Projektes ist es darüber hinaus, die komplexen und ambivalenten Erfahrungen und Zusammenhänge in einer Wissensbasis sammeln zu können. Dafür wird ein zentraler Aspekt dieses Teilprojektes die Entwicklung einer umfassenden Ontologie für Migrationsforschung/Migrationserfahrungen sein.

Mittels der in der Wissensbasis gesammelten, georeferenzierten und verknüpften Daten werden anschließend zusammen mit der Werkstatt des KompetenzwerkD Visualisierungsstrategien erarbeitet. Über interaktive Visualisierungen soll es möglich sein, neue Perspektiven über die Migrationsbewegungen und den dazugehörigen Erfahrungen zu gewinnen, die dann im gemeinsamen Webauftritt der Öffentlichkeit präsentiert werden.

2. Beschreibung der Daten und Metadaten

Projektdaten

Beschreibung: Zentrale Datengrundlage sind Interviews jüdischer Migrantinnen

(Ggf.) Ursprung der Daten: Austrian Heritage Collection des Leo Baeck Instituts <https://www.lbi.org/de/collections/austrian-heritage-collection/>

verwendete Datentypen und -formate: Die Interviews liegen als Audiodateien im mp3-Format zum Download bereit. Transkripte liegen nur in Ausnahmefällen vor.

Umfang: 157 mp3-Dateien mit ungefähr 220h Gesamtlänge

Dazugehörige Metadaten: Die Interviews sind im Repositorium des Leo Baeck Instituts mit typischen Metadaten angereichert, siehe beispielsweise: https://search.cjh.org/primo-explore/fulldisplay?docid=CJH_ALEPH000358318&context=L&vid=lbi&lang=en_US&search_scope=LBI&adaptor=Local%20Search%20Engine&tab=default_tab&query=any,contains,women%20refugees&sortby=rank&offset=160

Diese Metadaten werden teilweise übernommen.

Die Interviews sind offen mit einer Frage zur Lebensgeschichte gestaltet und unterscheiden sich sehr stark in dem, wie, was und in welchem Umfang die Interviewten auf einzelne Umstände der Flucht, ihren Erfahrungen mit Antisemitismus und ihrem Leben in der Emigration eingehen. Die Interviewpartner:innen vertiefen Aspekte jeweils den Umständen der Schilderungen entsprechend; beispielsweise in Fragen zum Familienhintergrund, der Beziehung zum Judentum oder dem Leben auf der Flucht.

Sollten sich weitere Datengrundlagen im Rahmen der Projektlaufzeit ergeben, werden diese ergänzt.

Wissensbasen bzw. Datensammlungen

Weedata

Beschreibung: Weedata¹ ist die zentrale im Projekt eingesetzte Wissensbasis. In ihr werden strukturierte Informationen zu Personen (Flüchtende), Interviews, Orten und Ereignissen zusammengetragen und verknüpft.

Datenmodell: Die Wissensbasis befindet sich derzeit im Aufbau. Zunächst werden Metadaten der Interviews und Personen sowie – falls vorhanden - strukturierte Daten zur Flucht (aus einer früheren Wissensbasis-Prototyp namens Semporalgraph) in die Datenbank importiert. Das Datenmodell wird später ergänzt.

Verwendete Vokabulare: Die in der Weedata-Wissensbasis verwendeten Vokabulare werden später ergänzt.

Referenzierte Normdatensätze: Weedata ermöglicht mittels integrierter Reconciliation-Dienste den Abgleich mit GND, HOV und Wikidata, deren IDs dann im Datensatz annotiert werden, z.B bei Orten oder bei dort verzeichneten Personen.

¹ <https://dikusa.saw-leipzig.de/toolbox/weedata/>

CATMA 7

Beschreibung: Das CATMA-System wird zur manuellen Annotation der Interview-Transkripte eingesetzt

Datenmodell: Es werden nur elementare Metadaten der Interviews in diesem System erfasst, Name der interviewten Person und Leo Baeck-ID des Interviews, um eine Zuordnung der Datensätze zu ermöglichen. CATMA dient nur der Erfassung von Annotationen. Die entsprechenden Annotations-ebenen und dabei verwendete Vokabulare werden im Folgenden erläutert.

Annotationsebenen / Verwendete Vokabulare: Gearbeitet wird bei der Annotation mit einer Reihe von Ebenen, die Details zur Flucht, der Emanzipation, sowie Erfahrungen mit Antisemitismus umfassen.

Aufgenommen wird eine Ebene mit den Stationen des Lebens, welche das Herkunftsland und -Stadt umfasst, sowie die Stationen der Flucht (Zwischenaufenthalte; gegebenenfalls mehrere aufeinander folgende Fluchtrouten) und den Lebensmittelpunkt bzw. die Punkte in der Emigration. Diesen wird jeweils ein Datum zugewiesen, wobei sich die Genauigkeit unterscheidet gemessen an der Nennung in den Interviews.

Eine weitere Annotationsebene umfasst die Emotionen insbesondere auf der Flucht. Hierbei werden Gefühle von Angst (um Familienangehörige oder der Unberechenbarkeit einer Situation), Trauer, Wut, Hoffnungslosigkeit, sowie Glück, Hoffnung, Aufregung, Liebe, Erleichterung, Sicherheit erfasst. Diesen wird auch jeweils eine Örtlichkeit sowie ein Datum zugewiesen und ist somit gekoppelt.

Für die kontextualisierte Darstellung der Daten gibt es eine Ebene mit historisch fixierten Ereignissen betreffend die Fluchtursachen und Möglichkeiten wie die Erlassung der Nürnberger Gesetze, die Reichsprogromnacht, den Anschluss Österreichs, den Überfall auf Polen oder die Einreiseverbote in einzelne Länder wie England oder die USA.

Daten zur persönlichen Emanzipation werden in einer Ebene aufgenommen, die Möglichkeiten, also die objektive Grundlage umfasst und die persönliche Realisation. Zum einen also die Möglichkeiten in Form von Bildung, Sprachenlehre und Ausbildung; und zum anderen den Werdegang der Interviewten beruflich, in Form von Beginn, Interruption durch Flucht oder andere Umstände und familiär, durch die temporäre oder dauerhafte Rolle als Hausfrau. Dazugehörig wird auch werden auch Momente der geäußerten Selbstrealisation aufgenommen, genauso wie Geäußertes zu politischem Engagement.

Die jeweilige Beziehung zum Judentum wird erfasst durch Angaben zum persönlichen Umfeld, der Praxis von Religion und der Kontakt und die mögliche Hilfe bei der Flucht durch Organisationen religiöser bzw. nicht religiöser Art.

Erste Erfahrungen mit Antisemitismus werden in einer Ebene erfasst, die sich untergliedert in Erfahrungen von Ausschluss, der Entrechtung, der Bedrohung und oder unmittelbaren Gewalt und der Expropriation von Eigentum. Auch diese Ebene wird gekoppelt angeführt gegebenenfalls mit den Ereignissen, sowie den Orten und Daten.

Eine Auflistung der beschriebenen Vokabulare findet sich im Anhang.

Referenzierte Normdatensätze: derzeit keine

3. Ethische und rechtliche Aspekte

Die zugrundeliegenden Daten der Austrian Heritage Collection des Leo Baeck Instituts sind für die Verwendung im wissenschaftlichen Kontext freigegeben. Laut Archiv ist das Ausfüllen einer allgemeinen Erlaubnisformulars² ausreichend um eine lizenzkonforme Verwendung und Veröffentli-

² <https://www.lbi.org/collections/reproductions-permissions-inquiries/>

chung zu ermöglichen. Dies wird zeitnah erfolgen, sobald weitere Details zur Zugänglichmachung der Projektergebnisse feststehen.

4. Speicherung, Archivierung und Sicherung der Daten (und Metadaten)

Arbeitsabläufe

Download / Webscraping: In einem ersten Schritt wurde die Datengrundlage beschafft. Dazu wurden die Audiodateien der Interviews im mp3-Format heruntergeladen. Außerdem wurden die dazugehörigen Webseiten geladen, mit dem Ziel, Metadaten zu den Interviews zu extrahieren.

Automatische Transkription: Es folgte eine Transkription der Interviews unter Verwendung von OpenAI Whisper³ (Version v20231117). Dabei wurde das Modell Large_v3 eingesetzt mit einem compression ratio threshold von 1,2, um fehlerhafte Stringwiederholungen möglichst zu verhindern. Die Ergebnisdateien wurden im json-Format ausgegeben und im Anschluss in das Eingabeformat von Spacy konvertiert.

Automatische Annotation: Unter Verwendung von spaCy⁴ folgte eine automatische Annotation der Interviewtexte. Zum einen wurde mit dem in spaCy enthaltenen NER-Modell eine Auszeichnung von Orten durchgeführt. Außerdem wurden auf Satzebene Emotionen annotiert. Dazu wurde auf Basis eines Huggingface-Datensatzes⁵ ein spaCy-Modell⁶ zur Emotionen-Klassifikation angelern und mit einem Konfidenzschwellwert von 0,7 angewendet.

Die Ergebnisse wurden in CATMA-kompatible Eingabedateien konvertiert (Rohtexte als txt, Annotationen als json).

CATMA ⁷ wird zur weiteren manuellen Annotation der Interviewtranskripte eingesetzt.

Annotiert werden daraufhin Textstellen, die einzelnen Ebenen zugewiesen und mit den jeweiligen Vokabularen versehen werden. Dies unterscheidet sich je nach Gegenstand und Umstand der Schilderung in der Länge, ob es sich beispielsweise um einen Fluchtpunkt handelt oder um eine Schilderung einer Situation von Entrechtung auf Grund von Antisemitismus während der Novemberprograme handelt. Dadurch werden Interviewsegmente genau referenzierbar für eine Darstellung und die Speicherung der Daten.

Die fertig annotierten Ergebnisse werden schließlich in die Weedata-Wissenbasis importiert. Die entsprechenden Workflows müssen noch ausgearbeitet und die Datenmodellierung abgeschlossen werden.

In Weedata erfolgt weitergehend der Aufbau einer Wissensbasis, die Informationen zu Personen (Flüchtenden), Interviews, Orten und Ereignissen enthält. Metadaten zu Interviews werden von der Online-Präsenz des Leo-Baeck-Instituts übernommen. Informationen zum Lebensweg, zur Fluchtroute usw. werden teils aus einem früheren Prototyp einer Wissensbasis übernommen, teils manuell in die Wissensbasis eingefügt.

Weitere Informationen werden später an dieser Stelle ergänzt.

³ <https://github.com/openai/whisper>

⁴ <https://spacy.io/>

⁵ <https://huggingface.co/datasets/dair-ai/emotion>

⁶ <https://heartbeat.comet.ml/emotion-classification-with-spacy-v3-comet-eaff310c0d7c>

⁷ <https://app.catma.de/seven/>

Datenspeicherung

Daten der Mitarbeitenden werden im HAIT Digilab gespeichert und sind dort vorerst nur passwortgeschützt zugänglich. Es ist nicht vorgesehen, die Daten zu löschen.

Die Ergebnisse und Zwischenergebnisse des Downloads, der Transkription und der automatischen Annotation werden für alle Projektbeteiligten zugänglich in der Nextcloud-Instanz der Sächsischen Akademie der Wissenschaften in einem Projektordner abgelegt.

CATMA 7 wird im Rahmen des kostenlosen Online-Services genutzt. Daten werden auf Servern der Universität Hamburg bzw. der TU Darmstadt gespeichert.

Weedata wird auf einem Server der Sächsischen Akademie der Wissenschaften gehostet.

Datensicherung

Backups der am HAIT gespeicherten Daten werden regelmäßig manuell über den Arbeitsrechner des Digilab-Administrators durchgeführt.

Von Daten, die an der Sächsischen Akademie der Wissenschaften gespeichert werden, werden entsprechend der hauseigenen Backupstrategie regelmäßig Sicherheitskopien angelegt.

Inhalte der CATMA 7-Online-Services werden ebenfalls regelmäßig durch den Anbieter gebackupt.

Archivierung

Die Daten werden als Teil des HAIT Digilab ohne Zeitbeschränkung archiviert. Löschungen sind nicht vorgesehen.

5. Datenverfügbarkeit und Dokumentation

Datenverfügbarkeit

Dem zentralen Ansatz des Verbundprojekts folgend erfolgt die Bereitstellung der zu veröffentlichen Datensätze zum einen in der Auszeichnungssprache RDF in einem noch festzulegenden üblichen Format (beispielsweise Turtle). Die in Weedata erfassten Daten werden dazu in einem automatisierten Export-Vorgang in das Schema der gemeinsamen Kernontologie⁸ überführt. Die exportierte RDF-Datei wird mittels eines öffentlich zugänglichen git-basierten Systems bereitgestellt, wodurch außerdem eine nachvollziehbare Versionierung der Datensätze gewährleistet wird.

Mindestziel ist die Bereitstellung der im Teilprojekt exportierten RDF-Daten in einem gut vernetzten Forschungsdatenrepositorium, um die Auffindbarkeit und Nachnutzbarkeit für Forschende zu verbessern. Hierfür wird das Repositorium RADAR4Culture von NFDI4Culture genutzt, welches bereits durch das KompetenzwerkD getestet wurde.

Die Bereitstellung weiterer im Teilprojekt anfallender Daten wird geprüft.

Eine Präsentation der Projektergebnisse in einem Webauftritt ist in Planung. Zentrum der Darstellung soll eine interaktive Karte sein, welche durch verschiedene thematische Einstiege in die Daten ergänzt wird.

⁸ <https://github.com/KompetenzwerkD/dikusa-core-ontology>

Dokumentation

Als Dokumentation der Daten an sich und der Arbeit mit den Daten soll dieser Datenmanagementplan dienen. Auf dieser Basis soll eine Nachnutzung der Projektdaten und Ergebnisse unterstützt werden.

Die Dokumentation der Datenrecherche liegt im Verantwortungsbereich der jeweiligen BearbeiterInnen.

6. Verantwortlichkeiten / Organisatorisches

Ansprechpartner für FDM im HAIT-Projekt ist Chris Weinhold.

Ansprechpartner für FDM im DIKUSA-Projekt ist Dirk Goldhahn.

Kostenfragen

Die Serverkosten für das HAIT Digilab betragen ca. 70 Euro monatlich und sind über Projektmittel gedeckt.

Personalkosten am HAIT sind ebenfalls über Projektmittel gedeckt.

Weitere Personalkosten an der SAW Leipzig sind über Haushaltsmittel gedeckt (Personalkosten KompetenzwerkD, Serveranschaffung bzw. Miete)

7. Anhang

Verwendete Vokabulare:

Stationen des Lebens: Eine Zuordnung zu Orten und Zeiträumen erfolgt darauf aufbauend in der Wissensbasis Weedata

point of origin1

escape route1.1

stopover1.1

escape route1.2

stopover1.2

escape route1.3

point of arrival1

point of origin2

...

point of arrival2

place of birth

place of death

Emotionen

excitement

anger

exclusion

fear

 fear for family members

 fear about incalculability

hope

hopelessness

hunger

incomprehension

joy
love
relief
sadness
safety
surprise

Ereignisse

01.10.1938 Annektierung des Sudetenlandes
01.09.1939 Überfall auf Polen
03.09.1939 Einreisesperre nach England
09.11.1938 "Reichskristallnacht"
10.05.1940 Besetzung Luxemburgs
12.03.1938 Anschluss Österreichs
15.03.1939 Errichtung des Protektorats Böhmen und Mähren
15.09.1935 Reichsbürgergesetz/ Nürnberger Gesetze
1924-1948 Johnson-Reed Act/ starke Reglementierung der Einwanderung durch Quoten nach Nationalitäten
1933-1937 (formal - 1941) Ha'avara-Abkommen
1938-1941 International Settlement Shanghai
07.1941-1948 nahezu vollständige Unmöglichkeit der einreise in die USA
10.1939-04.1940 Einreiseverbot für das britische Mandatsgebiet Palästina
10.1941 Verbot der Ausreise für Juden aus NS-kontrollierten Gebieten

Erfahrungen mit Antisemitismus

expropriation
exclusion
attack
fear of attacks
disenfranchisement

persönliche Emanzipation

opportunities for emancipation

study

education

language learning

apprenticeship

career

transition

termination

beginning

change of role

housewife

interruption

political commitment

party mandate

NGO

feminist

self-realization